Image captioning in Turkish language: Database and model

Image captioning is the art of describing an image with sentences. The explanation of an image requires several tasks including the recognition of salient objects present in the picture, the understanding of their semantic relationships, the comprehension of the scene represented in the background and the capability to convert that knowledge into a syntactically correct sentence. A practical application of this research is also to support blind people with description of the surrounding environment. To date, the number of studies in Turkish language is still too limited and requires further investigation

This work used Python code to automatize Yandex translation API and convert all captions of the MS COCO (Lin et al. [20]) database from English language into Turkish language. The resulting Turkish captioned MS COCO database was used to test the proposed model for image captioning in Turkish language. Considering the recent developments in the machine translation field, the used image-captioning model employs an encoder-decoder framework, where a Convolutional Neural Network (CNN) encodes the image into a fixedlength vector representation, and a Long-Short Term Memory (LSTM) maps those vectors and generates image descriptions in Turkish language. In this study, we created two models. In Model-1 the weights of the used pre-trained CNN were frozen, while in Model-2 CNN and LSTM were fine-tuned together.

The proposed models were evaluated using both human based evaluations, and the most common metrics such as BLEU, METEOR, ROUGE and CIDEr. Both qualitative and quantitatively evaluations were satisfactory. In all cases, Model-2 had higher performance.

This study introduces a novel Turkish captioned database together with a model to generate captions in Turkish language. The provided Web application will allow for crowd sourcing and the resulting Turkish captioned MS COCO database will be available for research purpose